

## A HuComTech audio adatbázis szintaktikai szintjének elvei és szabályrendszerének újdonságai

Kiss Hermina<sup>1</sup>

HuComTech Group, Debreceni Egyetem, Általános és Alkalmazott Nyelvészeti Tanszék,  
4032 Debrecen, Egyetem tér 1.  
kissh3@gmail.com

**Kivonat:** A HuComTech multimodális adatbázis egyik annotációs szintje a szintaktikai szint. Az annotációs szempontrendszer kialakítása során újbóli átgondolásra került a mondat fogalma, a tagmondatok hierarchiájának elemzési módszere és az implicit nyelvi elemek kimutatásának módszertana. Ennek tükrében létrehoztunk egy új típusú mondatelemzési módszert, aminek szintaktikai alapegysége a tagmondat. Az adatbázis legfontosabb alapelvei: az adatbázis legyen preteoretikus, tükrözze a különböző tudományos megközelítések közötti konszenzust, valamint legyen aluspecifikált. A spontánbeszéd-kutatás szintaktikai elemzésének speciális jellegét azzal lehet leginkább kiemelni, ha különös figyelmet fordítunk az implicit nyelvi elemek összegyűjtésére és rendszerezésére, valamint a tagmondatok hierarchiájának jellemzésére. Ez az előadás erre vállalkozik.

### 1 Bevezetés

A munkánk alapját a HuComTech spontánbeszéd-korpusz és adatbázis képezi. Az adatbázis a kommunikáció számos multimodális jegye mellett nyelvi, ezen belül a beszédre is vonatkozó adatot tartalmaz. Külön kihívás a folyamatában megszülető, a kommunikáció során kialakuló spontánbeszéd mondatnyi elemzése, hiszen az gyakran ellenáll a hagyományos mondatelemzésnek. Elemzése és annotálása számos problémát vet föl egyrészt azért, mert a beszélő még nem tudja, hogy az általa kifejezendő információ milyen szerkezetben fog megjelenni, másrészt pedig az élőszó spontaneitásának gyakori következménye a pongyola nyelvhasználat, ami egy nem kellőképpen átgondolt és nem megfelelően létrehozott szintaxist hoz létre. Első és legfontosabb dolog a spontán beszéd annotációs szabályainak kialakításához, hogy meghatározzuk a használandó alapfogalmakat. Mivel jelen esetben két személy közötti kommunikáció szintaktikai elemzéséről van szó, minden esetben az egyes beszélők által megvalósított egyes fordulókat tekintjük az elemzés tárgyának. Az egyes fordulókon belül azonosítjuk a szintaktikai struktúrát. A szintaktikai struktúra alapjának a tagmondatot tekintjük (mélyebb bontásra már csak azért sem vállalkozunk, mert ezt a beszélt nyelvi produkció gyakran nem is teszi lehetővé) és ezt szerkezeti sajátosságai alapján határozzuk meg. Az elemzés és az annotáció

---

<sup>1</sup> A jelen tanulmány alapjául szolgáló kutatásban a szerzőt *A felsőoktatás minőségének javítása a kutatás-fejlesztés-innováció-oktatás fejlesztésén keresztül a Debreceni Egyetemen* című, TÁMOP-4.2.1/B-09/1/KONV-2010-0007 projektazonosítójú program támogatta.

egységes strukturális szempontok alapján azt ígéri, hogy az elemzés jól tükrözi a nyelv beszédben kialakuló szerkesztését, ugyanakkor kellően alulspecifikált ahhoz, hogy különböző elméleti megközelítésekben is jól használható legyen.

## 2 A mondat és a tagmondat fogalma

A mondat fogalmának definícióját olyan szempontból közelítjük meg, hogy érvényesüljön az az alapvető célunk, miszerint az általunk kidolgozott szintaktikai modell preteoretikusan működtethető, tehát az ember-gép közötti kommunikáció tanulmányozására létrejött adatbázisban szinkronba hozható a nyelvészeti szakirodalom mondatfogalmának többféle szempontú megközelítése és ennek megfelelően többféle meghatározása.

Ennek a célnak az egyik velejárója az, hogy elemzési szempontrendszerünk alulspecifikált, hiszen a tagmondatok közötti viszonyok meghatározása után nem bontjuk tovább az elemzési szempontrendszert úgy, hogy az alá- és mellérendelő mondat típusok megnevezését is lehetővé tegyük.

A *Strukturális magyar nyelvtan* mondattanról szóló kötetében az alárendelő mellékmondat vonzatnak minősül, ezért nem érvényes az a szerkezeti meghatározás, miszerint a mondat szerkezete egyszerű és összetett mondatokból áll össze [1]. Mi viszont elfogadjuk, hogy a kifejtett mondatrész külön tagmondat, hogy minél részletesebben és érzékletesebben kimutassuk a mondat implicit elemeit. Nem mondjuk azt tehát, hogy az alárendelt tagmondat egy vonzat, és nem hiányzik semmi a mondatból, hanem külön tagmondatként értelmezve felszínre hozzuk az így kimutatható implicit nyelvi elemeket.

Ebbe a rendszerbe beilleszthető a vonzatról való felfogásunk, amit a *Strukturális magyar nyelvtan*, illetve a *Magyar grammatika* [2] is elfogad: vonzatnak az elhagyhatatlan bővítményeket tartjuk, ami azt jelenti, hogy a vonzat a grammatikai struktúra sérülése nélkül nem hagyható el a nyelvi egység mellől, amihez tartozik. Az alanyt viszont nem tekintjük vonzatnak.

Ennek megfelelően az elemzésünk alapegysége a tagmondat. A tagmondat szerkezetileg nem más, mint szavak kapcsolódása egy hierarchikus rendben. Egy tagmondat szerkezeti határát az képezi, amikor egy adott szót már nem tudunk az addig (az azt lineárisan megelőző és/vagy követő szavakból) felépült hierarchikus rendben elhelyezni. Funkcionálisan egy hiánytalan tagmondat a régensből (állítmány) és kötelező vonzataiból, valamint az alanyból áll. Számunkra az állítmány az igét és annak vonzatait jelenti együttesen, tehát nem csupán a leíró nyelvtan szerinti egyszerű és összetett állítmányt, hanem azzal együtt a vonzatokat is magába foglalja.

A szerkezetek láncszerű grammatikai kapcsolata tagmondatok sorát alkotja meg. Ezek, ha szerkezetileg kapcsolódnak, mondatná állnak össze. A mondat tehát a tagmondatok láncszerű, szerkezeti kapcsolódása és minimum egy tagmondatból áll.

### 3 Implicit nyelvi elemek

A beszélt nyelvben gyakori elemek az ismétlések, a töltelékszavak, a mondatok megszerkesztettsége lazább, szabálytalanabb. Ennek egyik grammatikai következménye az, hogy elmaradhat a főmondat, az utalószó, a kötőszó, a grammatikai, illetve logikai alany, az állítmány, a tárgy, a jelző és az ige. Ezen nem jelölt nyelvtani elemekre bizonyos esetekben következtethetünk akár strukturálisan, akár szemantikailag/kontextuálisan, más esetekben azonban nem (pl. a megkezdett, de befejezetlen tagmondatok esetében). A grammatikailag jólformált és nem jólformált tagmondatokat egyazon szempontrendszer alapján elemezzük.

### 4 Minimális mondat

A beszélt nyelv lazább szerkesztettségének fentebb bemutatott grammatikai következménye az implicit nyelvi elemek gyakori előfordulása mellett egy másik fontos grammatikai következménye az, hogy egy-egy forduló [3] állhat különálló szavak olyan egymásutániságából, amelyek között semmilyen grammatikai összerendeződés nincs. A tagmondat fentebbi meghatározása alapján ilyen esetekben ezen szavak külön-külön egyetlen tagmondatból álló mondatokat képeznek. Ezek a minimális mondat esetei. Külön figyelmet kell fordítanunk a lexikális tartalom nélküli hangzó megnyilvánulásokra. Ezek a lexikális tartalom nélküli minimális mondat esetei. Csak azokat az eseteket vesszük figyelembe, amelyek a fordulók elején vagy végén jelennek meg. (A tagmondat szavai között megfigyelhető, gyakran bizonytalanságot vagy a kifejezendő gondolat módosítását jelző hangzó megnyilvánulásokat, mint amik nem befolyásolják a mondatszerkezetet, nem jelöljük.) A minimális tagmondatra a példák a következő alfejezetek:

#### 4.1 A befejezetlen tagmondatok

például: *De... És... Hm ... Úúú...*

#### 4.2 A mondatszók

például: köszönések, megszólítások, indulatszavak, töltelékszavak, stb.

#### 4.3 Egyszavas válaszok

például: *Igen. Nem. Talán.*

#### 4.4. Egymondatos visszakerdezések

például: *ugye*, (akár visszakerdezés, akár töltelékszó), *Legjobb főnök? Legszebb élmény?*

De az *így/ügy* töltelékszavakat nem soroljuk ide, mivel grammatikailag (határozóként) kapcsolódnak a tagmondathoz.

#### 4.5. A mint-tel kezdődő hasonlító határozói alárendelt tagmondatok

Olyan lettem, *mint te*. Ez már nem olyan, *mint az volt*.

#### 4.6. Töltelékszavak

*ugye, így/ügy tehát*, stb.

### 5 Teljes tagmondat

Fentebb a beszélt nyelvre különösen jellemző, valamilyen szempontból hiányos szerkezetekről szoltunk. Természetesen a beszélt nyelvben is találkozunk az ettől különböző szerkesztéssel, azaz a strukturális szempontból teljes mondatokkal. Ezek funkcionálisan tartalmazzák a régenst (állítmányt) annak kötelező vonzataival és az alanyt. Ezt a leíró nyelvtan egyszerű mondatának nevezhetjük, illetve összetett mondat esetén az összetételeket.

### 6 A tagmondatok kódolása

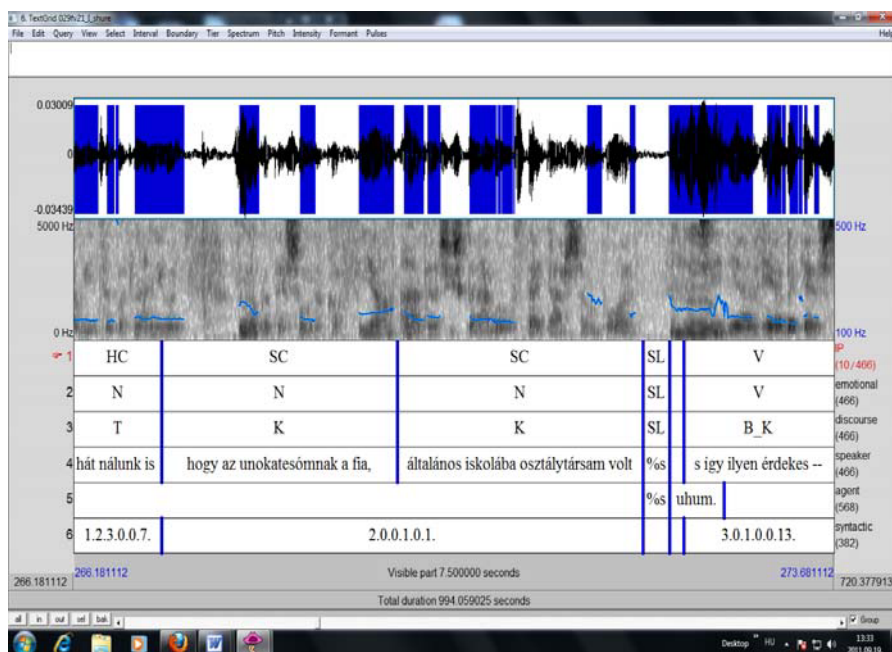
A tagmondatok láncolata lineárisan és hierarchikusan is szervezi a beszédet. Ennek feltárása alapvető célunk. Ennek megfelelően meghatározunk az alá- és mellérendelő tagmondatokat, illetve a tagmondatok közötti grammatikai kapcsolat hiányát (beágyazást, beékelést).

#### 6.1 Szegmentációs szakaszok

Alárendelő tagmondatok esetén egyértelmű a szegmentációs határhelyzet, azaz a tagmondat határa. Mellérendelő tagmondatok esetén vagy új mondat indul kötőszóval, illetve anélkül kezdve, vagy az előtte lévő tagmondathoz kapcsolódik, s így még ugyanannak a tagmondatfüzérnek a tagja, amihez az előző kapcsolódik.

## 6.2 A számozás

A számozás a tagmondatok közötti sorrendiséget és a tagmondatok közötti viszonyt fejezi ki. A számozás kezdete a hagyományos mondat kezdetét jelöli. A számozás ott fejeződik be, ahol a hagyományos mondat végét lehet érzékelni. A hagyományos mondat végét nem az intonáció és elsődlegesen nem a szemantika, illetve interpretáció határozza meg, hanem a szintaktika.



1. ábra: A szintaktikai annotációs szint kódolása.

Az 1. ábrán láthatjuk a kódolási rendszert, az annotáció 6. szintjén. A kódrendszerben az első szám tehát a tagmondatok sorszámát jelenti. A második szám azt jelöli, hogy az adott tagmondathoz tartozik-e alárendelés, és ha igen, akkor hányas számú tagmondat. Ha nincs, akkor az 0 értékkel van jelölve. A harmadik szám a tagmondathoz tartozó mellérendelő tagmondat(ok) sorszámát jelöli. Ha nincs ilyen, akkor a 0 érték látható. A negyedik számjegy azt mutatja meg, hogy az adott tagmondat hányas számú tagmondathoz tartozik az alárendeltje. Itt is megjelenhet a 0 érték. Az ötödik számjegy a grammatikai kapcsolat hiányát jelöli, azt mutatja meg, hogy melyik tagmondat kapcsolódik hozzá úgy, hogy grammatikai elem nem jelenik meg. A számok között pont van. Ha egy elemzési szemponthoz több szám is tartozik, akkor azok vesszővel vannak elválasztva.

## 7 A hiány kategóriái

### 7.1. Nem hiányzik semmi

Nem hiányzik semmi abban az esetben, ha érvényesül a teljes tagmondat fent leírt definíciója.

### 7.2 Hiányzik a főmondat

*Mert szeretnék munkát.*

*Ha így lesz.*

*Mikor még kicsi voltam.*

### 7.3 Hiányzik az előtte álló mellékmondat

Abban az esetben használjuk ezt a kategóriát, amikor a tagmondat előtt nincs tagmondat, (az előző mondathoz tartozó tagmondat).

*És ő nem vette fel a telefont.*

*Meg el sem jött*

*De én mindenképp el akartam menni.*

### 7.4 Hiányzik a kötőszó

*Attól függ, mit nézünk.*

*Éreztem, hogy pályakezdőként itt sokat tanulhatok.*

*Emlékszem, amikor ezt tavaly átéltem.*

### 7.5 Hiányzik az utalószó

*Angolt tanultam úgy eddig is, mert nekem az egyetemen kellett.*

*Sokszor dolgoztam már, hogy minél tapasztaltabb legyek.*

*Ne mondjátok meg, hogy hová kell menni.*

### 7.6 Hiányzik a grammatikai alany

*Csak úgy nem ilyenre számítottam.*

*Megyek dolgozni.*

*Nagyon fontos dolgokat mondott nekünk.*

### 7.7 Hiányzik a logikai alany

*Hát, általában így szokott lenni.  
Nincs szükségem erre egyáltalán.  
Nem volt még előző munkahelyem.*

### 7.8 Hiányzik az állítmány

Például: *A főnök kabátban.* Abban az esetben hiányzik az állítmány, ha az ige és annak vonzatköre nem jelenik meg a tagmondatban, van(nak) viszont egyéb szabad határozó(k).

### 7.9 Hiányzik a tárgy

*Sokszor iszik valóban.  
Akkor így nem vették észre.  
Ő is látta.*

### 7.10 Hiányzik a határozó

Például: *a megy* ige vonzatai: vki, vhová. Ha ezek közül hiányzik a határozó, akkor az hiánynak van feltüntetve.

*Nem hitt.  
Részt vett.  
Pista jártas.*

### 7.11 Hiányzik a jelző

*Liter tejet hozott.  
Köbméter víz fogyott.  
Kiló kenyérrel tért vissza a munkahelyére.*

### 7.12 Hiányzik az ige

Például: *János spagettit.* Ha a tagmondatban megvan(nak) a kötelező vonzat(ok), de a régens hiányzik (eszik/evett/ fog enni).

*Péter a kávé.  
János könyvet.  
A lisztet.*

### 7.13 Befejezetlen tagmondat

Az élőbeszédre jellemző sajátosság, hogy a nyelvtani korrekciók a beszéd folyamatában történnek meg. Ennek grammatikai következménye az, hogy a szerkesztés befejezetlen marad. A tagmondat meghatározás alapján azonban az ilyen befejezetlen szerkezeteket tagmondat értékűnek tekintjük. A befejezetlenséget azonban külön kódoljuk, ugyanis feltesszük, hogy a befejezetlenség által keltett információhiányt egy másik, nem nyelvi modalitás pótolja és így az azonosítható pl. egy arckifejezésben vagy egy mozdulatban stb. Így a szintaktikai annotálás mint a multimodális annotálás része hozzájárulhat ahhoz, hogy az egyik modalitásból hiányzó elemet egy másik modalitás ugyanazon időpillanatában kutathassuk, tehát egy befejezetlen mondat kézmozdulatokkal, mimikával való lezárását nyomon követhessük a szintaktikai szinten is.

### 7.14 A hiány nem releváns

A hiány nem releváns akkor, ha nem tudunk érvényes hiány kategóriát megállapítani, de a mondat mégsem tekinthető teljesnek.

#### 7.14.1 Mondatszók

##### 7.14.1.1 Indulatszavak:

*Hú! Nahát! Ó! stb.*

##### 7.14.1.2 Igenlő egyszavas válaszok:

*De! Igen! Rendben! Jó! stb.*

##### 7.14.1.3 Tagadó egyszavas válaszok:

Nem. Módosítószóval együtt: Még nem. Szerencsére nem. Nem nagyon. Én nem.  
*Innen még nem.*

De ha a tagadószó mondatrészben van, akkor az alany és az állítmány hiányzik: *Azt mondom, hogy nem. Utazáshoz tudnám kötni, de igazából még nem.* stb.

##### 7.14.1.4 Bizonytalan egyszavas válaszok:

*Talán. Lehet. Bizonyára.* stb.



#### **7.14.1.5 Köszönések:**

*Viszlát! Viszontlátásra! Jó napot! De a Jó napot kívánok!* köszönésforma nem tartozik ehhez a kategóriához, mert egyértelműen meg tudjuk határozni a mondatban az alanyt, az állítmányt és a vonzatot.

#### **7.14.1.6 Udvariassági formulák:**

*Szívesen! Nagyon szívesen!* stb.

#### **7.14.1.7 Töltelékszavak**

*Hát, ugye, így, úgy,* stb.

#### **7.14.1.8 Megszólítások**

*András! Kinga!* stb.

#### **7.14.2 Egymondatos visszakérdezések**

például: *ugye?* (akár visszakérdezés, akár töltelékszó), *Legjobb főnök? Legszebb élmény?* De az *így/úgy* töltelékszavakat nem soroljuk ide, mivel grammatikailag (határozóként) grammatikailag kapcsolódnak a tagmondathoz.

#### **7.14.3 Mint-tel kezdődő hasonlító határozói alárendelt tagmondat esetén**

*Szebb, mint az.*

*Sokkal jobb lesz így, mint úgy.* stb.

#### **7.14.4 Valamilyen okból (például a pongyola nyelvhasználat mértéke miatt) kikövetkeztethetetlen tagmondatok esetén**

*Ha véletlenül találkozunk egy szíát, de több nem.*

## **8 Összegzés**

A Praat szoftver felhasználásával olyan annotációs szabályrendszert dolgoztunk ki, amely lehetővé teszi a spontán beszéd szintaxisának kutatását. Különös hangsúlyt fektettünk arra, hogy a spontán beszéd jellegzetességeit kezelhetővé tegyük a magyar

nyelv mondattana keretei között, mint ami rendszerében nem, csak megvalósulásában különbözik attól. Nem tettünk említést számos problémakörrel, amelyek az adott kategóriák átgondolását segítették. Például az egyedi szó- és nyelvhasználatból adódó jelenségekről, sajátosságokról, vagy a töltelékszavak, indulatszavak spontán beszédbe illeszkedő rendszeréről, illetve a pongyola nyelvhasználat következményeként létrejövő szintaktikai problémákról. (Mint például az abszolút és relatív főnév elhelyezkedése a mondat hierarchiájában, a kötőszóval kezdődő mondatok kérdéséről, a főnevesült jelző mondattani szerepköréről, a függő beszédben jelen lévő implicit elemekről, az ellipsis számos kérdésköréről, illetőleg a dialógus másik szereplőjének a vizsgált személy grammatikájára tett hatásáról.) Itt ismertetett kódrendszerünk lehetővé teszi azt, hogy az adatbázist vizsgáló kutatók további szintaktikai elemzéseket folytassanak, kiegészítve, részletezve az általunk létrejött rendszert.

## Bibliográfia

1. Keszler B.: Szintagmatan. In: Keszler B. (szerk.): Magyar Grammatika. Nemzeti Tankönyvkiadó, Budapest (2000) 355
2. Komlósy A.: Régenek és vonzatok. In: Kiefer F. (szerk.): Strukturális Magyar Nyelvtan I. Akadémiai Kiadó, Budapest (1992) 308—316
3. Iványi Zs.: A nyelvészeti konverzációelemzés. Magyar Nyelvőr Vol. 125 (2001) 74-93  
[<http://www.c3.hu/~nyelvor/period/1251/125106.htm>]